A Mini Project Synopsis on

Driviction(Car Price Prediction)

S.E. – Computer Science and Engineering-Data Science

Submitted By

Vedant Parulekar	**21107034**
Tanaya Patil	**21107017**
Harsh Mulik	**21107044**
Ridhvik Thakur	**21107056**

Under The Guidance Of

Prof. Rajashri Chaudhari

DEPARTMENT OF CSE-DATA SCIENCE

A.P. SHAH INSTITUTE OF TECHNOLOGY

G.B. Road, Kasarvadavali, Thane (W), Mumbai-400615

UNIVERSITY OF MUMBAI

Academic year: 2022-23

CERTIFICATE

This to certify that the Mini Project report on Driviction(Car Price Prediction) has been submitted by Vedant Parulekar 21107034, Harsh Mulik 21107044, Ridhvik Thakur 21107056, Tanaya Patil 21107017 who are a Bonafede students of A. P. Shah Institute of Technology, Thane, Mumbai, as a partial fulfilment of the requirement for the degree in Computer Science and Engineering (Data Science), during the academic year 2022-2023 in the satisfactory manner as per the curriculum laid down by University of Mumbai.

Prof. Rajashri Chaudhari

Guide

Prof. Anagha Aher — Dr. Uttam D. Kolekar

Head Department of CSE-Data Science — Principal

External Examiner(s)

1.

2.

Place: A. P. Shah Institute of Technology, Thane

ACKNOWLEDGMENT

This project would not have come to fruition without the invaluable help of our guide Prof. Rajashri Chaudhari. Expressing gratitude towards our HOD, Prof. Anagha Aher, and the Department of Data Science for providing us with the opportunity as well as the support required to pursue this project. We would also like to thank our teacher Prof. Kiran Babar who gave us her valuable suggestions and ideas when we were in need of them. We would also like to thank our peers for their helpful suggestions.

TABLE OF CONTENTS

Chapter 1

Introduction

The pre-owned vehicle market is a developing business with a market esteem that has almost multiplied itself in earlier years. The ascent of online sites and other instruments like it has made it more straightforward for the two purchasers and merchants to improve comprehension of the variables that decide the market worth of a preowned vehicle. In light of a set of variables, Machine Learning calculations might be used to conjecture the cost of any vehicle. The informational collection will remember data for an assortment of vehicles. There will be data with respect to the vehicle's specialized components, for example, the motor kind, fuel type, total driven kilo meters, and that's only the tip of the iceberg, for each vehicle. There is no inclusive instrument for building up the retail cost of utilized vehicles in light of the fact that unique sites utilize various techniques to make it. By utilizing measurable models to expect to value, it is conceivable to acquire a fundamental value gauge without entering every one of the subtleties into the ideal site. The fundamental motivation behind this study is to think about the precision of two distinct expectation models for assessing a pre-owned vehicle's retail cost. Subsequently, we offer a Machine Learning-based philosophy at anticipating the costs of second-hand vehicles in light of their attributes. This philosophy can help purchasers hoping to buy a pre-owned vehicle in making more informed decisions. Clients can now search for all vehicles in a district without actual endeavors, whenever and from any area. With the Coronavirus sway on the lookout, we have seen lot of changes in the vehicle market. Presently some vehicles are sought after subsequently making them exorbitant and some are not popular consequently less expensive. With the adjustment of market due to Coronavirus 19 effect, people/sellers are facing issues with their past Car Price valuation AI/Machine Learning models. Along these lines, they are searching for new AI models from new information. Here we are building the new car price valuation model. The primary point of this venture is to create a dataset with the help of web scraping and anticipate the cost of trade-in vehicle in view of different elements. The costs of new vehicles in the business are fixed by the producer for certain extra expenses brought about by the Government as assessments. Along these lines, clients purchasing another vehicle can be guaranteed of the money/investment they contribute to be commendable. Be that as it may, because of the expanded cost of new vehicles and the ineptitude of clients to purchase new vehicles because of the absence of assets, utilized vehicles deals are on a worldwide increment. There is a requirement at a pre-owned vehicle cost expectation framework to successfully decide the

value of the vehicle utilizing an assortment of highlights. Despite the fact that there are sites that offers this assistance, their expectation technique may not be awesome. Additionally, various models and frameworks might contribute on anticipating power for a preowned vehicle's genuine market esteem. It is essential to realize their genuine market esteem while both trading. To have the option to anticipate utilized vehicles market worth can help the two purchasers and merchants. Utilized Vehicle merchants are one of the greatest objectives gathering that can be keen on consequences of this review. On the off chance that pre-owned vehicle merchants better get what makes a vehicle attractive, what the significant highlights are for a pre-owned vehicle, then, at that point, they might think about this information and proposition a superior assistance.

1.1 Purpose

There are a few major worldwide multinational participants in the automobile sector, as well as a number of merchants. By trade, international companies are mostly manufacturers, although the retail industry includes both new and used automobile dealers. The used automobile market has seen a huge increase in value, resulting in a bigger percentage of the entire market. In India, about 3.4 million automobiles are sold each year on the second-hand car market. Using well-known algorithms from Python libraries, we were able to successfully construct machine learning algorithmic paradigms. On our dataset, we first do pre-processing and data cleaning. We trimmed the tuples that contained null values, which accounted for less than 1% of the total. The findings revealed a positive relationship between price and kilometers travelled, as well as year of registration and kilometers travelled. Negative correlation is related to the notion of inverse proportion, whereas positive correlation is related to the concept of direct proportion. The model was trained using several tuples.

1.2 Objectives

1. Data Collection: To scrape the data of several used cars from various websites like OLX, Car Dekho, Cars24 etc.

2. Model Building: To build a supervised machine learning model for forecasting value of a vehicle based on multiple attributes.

3. To build a machine learning model that can accurately predict the price of a used car based on various factors such as age, mileage, model, condition, and location.

4. To provide accurate predictions of used car prices, the system can assist buyers in making informed decisions about whether to purchase a particular vehicle or not.

5. To provide extra and intricate features of the car, the customer is interested in.

Chapter 2

Problem Definition

The main aim of this project is to predict the price of used cars using the various Machine Learning (ML) models. This can enable the customers to make decisions based on different inputs or factors namely

• Brand or Type of the car one prefers like Ford, Hyundai.

• Model of the car namely Ford Figo, Hyundai Creta.

• Location like Delhi, Chennai, Mumbai.

• Year of manufacturing like 2020, 2021.

• Type of fuel namely Petrol, Diesel.

• Price range or Budget.

• Type of transmission which the customer prefers like Automatic or Manual.

• Mileage

to name a few characteristic features required by the customer. The project Car Price Prediction deals with providing the solution to these problems. Through this project, we will get to know which of the factors are significant and tell us how they affect the car's worth in the market.

Steps Involved:

1. Data Cleaning which involves identifying all the null and missing values and removing the outliers.
2. The next process is Data Preprocessing that involves us to normalize and standardize the dataset.
3. We then use the different Machine Learning models like Linear Regression, XGBoost.
4. We then need to compare the performance of any of the models used.
5. Now, we gather insights and analyze the data based on the accuracy test.

Chapter 3

Proposed System

There are a few major worldwide multinational participants in the automobile sector, as well as a number of merchants. By trade, international companies are mostly manufacturers, although the retail industry includes both new and used automobile dealers. The used automobile market has seen a huge increase in value, resulting in a bigger percentage of the entire market. In India, about 3.4 million automobiles are sold each year on the secondhand car market.

3.1 Features and Functionalities

A car price prediction machine learning model can have various features and functionalities, some of which are:

Data collection: The model requires a large dataset consisting of historical car prices, as well as other features such as car make and model, year, mileage, location, and other relevant data points.

Data preprocessing: The data collected needs to be cleaned, formatted, and prepared for analysis. This may include removing duplicate data, filling in missing values, and converting categorical variables into numerical representations.

Feature selection: The model needs to identify which features are most important in predicting car prices. This can be done using techniques such as correlation analysis or feature importance ranking.

Model selection: There are various machine learning algorithms that can be used for car price prediction, including linear regression, decision trees, random forests, and neural networks. The model should be selected based on its accuracy, computational requirements, and other relevant factors.

Model training: The selected model is trained on the preprocessed data, using techniques such as cross-validation to optimize its performance.

Hyperparameter tuning: The model's hyperparameters, such as learning rate and regularization

strength, are tuned to further improve its performance.

Model evaluation: The model's accuracy is evaluated using metrics such as mean squared error, mean absolute error, and R-squared.

Deployment: The final model can be deployed as a web application or API, allowing users to input car information and receive a predicted price.

Continuous learning: The model can be updated with new data as it becomes available, allowing it to continuously improve its predictions over time.

Overall, a successful car price prediction machine learning model should have robust data collection and preprocessing capabilities, accurate and efficient modeling algorithms, and effective evaluation and deployment strategies.

Chapter 4

Project Outcomes

A car price prediction machine learning model can provide various project outcomes, such as:

1. Predictive accuracy: The model can accurately predict car prices based on various factors such as car make and model, year, mileage, location, and other relevant data points. This can help car buyers and sellers make informed decisions about pricing and purchasing.
2. Increased efficiency: By automating the process of price prediction, the model can save time and resources that would otherwise be spent manually researching and analyzing car prices.
3. Improved transparency: The model can provide transparency into how car prices are determined, making the pricing process more objective and less susceptible to bias.
4. Enhanced customer experience: Car buyers can use the model to research and compare prices across different models and locations, helping them find the best deals and make informed decisions.
5. Increased revenue: Car dealerships and other businesses can use the model to optimize their pricing strategies, potentially increasing revenue by selling cars at optimal prices.
6. Insights into market trends: The model can provide insights into market trends and demand for certain car models, helping businesses make strategic decisions about inventory and pricing.

Overall, a successful car price prediction machine learning model can provide numerous benefits to both individuals and businesses in the automotive industry, including improved accuracy, efficiency, transparency, and customer experience.

Chapter 5

Software Requirements

Libraries for Machine Learning Models:

```
#libraries for preprocessing

from sklearn import preprocessing

from sklearn.preprocessing import StandardScaler

from sklearn.preprocessing import MinMaxScaler



#libraries for evaluation

from sklearn.metrics import mean_squared_log_error,r2_score,mean_squared_error

from sklearn.model_selection import train_test_split



#libraries for models

from sklearn.linear_model import LinearRegression

from sklearn.linear_model import Ridge

from sklearn.linear_model import LassoCV,RidgeCV

from yellowbrick.regressor import AlphaSelection

from sklearn.linear_model import Lasso

from sklearn.neighbors import KNeighborsRegressor

from sklearn.ensemble import RandomForestRegressor
```

```
from sklearn.ensemble import BaggingRegressor

from sklearn.tree import DecisionTreeRegressor

from sklearn.svm import SVR

from sklearn.ensemble import AdaBoostRegressor



#Loading Dataframe

#df=pd.read_csv("vehiclesFinal2.csv")

df=pd.read_csv("C:/Users/Vedant/OneDrive/Desktop/vehicles_final.csv")

#df=df.drop('Unnamed: 0',axis=1)

df=df.drop('id',axis=1)

df2=df.copy()

df.head()
```

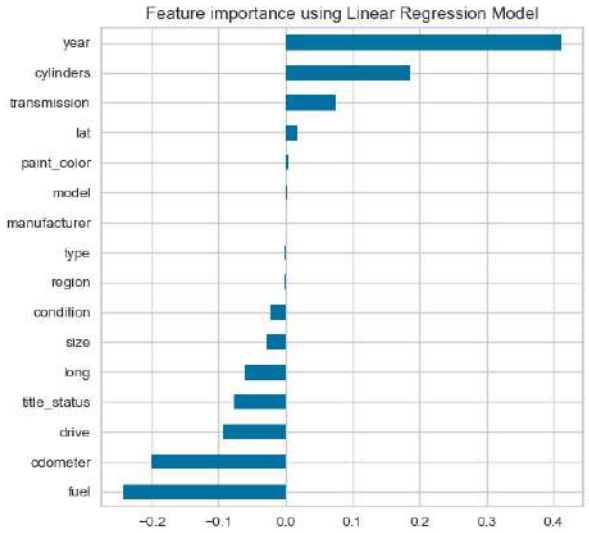
Steps:

1) Transforming categorical columns values into integer values
2) Scaling Dataset
3) Model Implementation

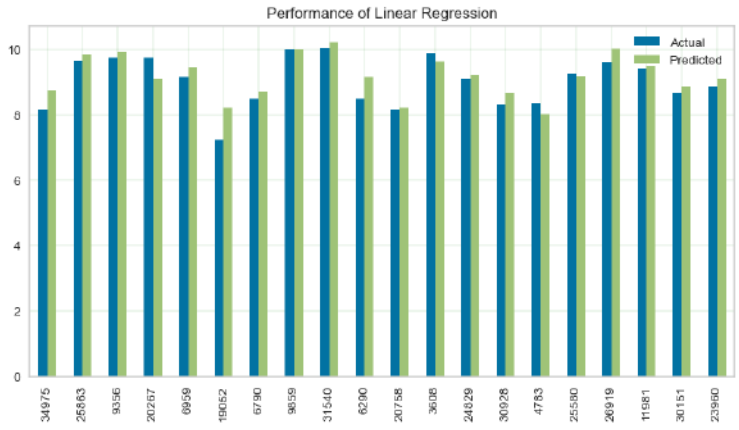
    - Linear Regression
    - Ridge Regression
    - Lasso Regression
    - K-Neighbors Regressor
    - Random Forest Regressor
    - Bagging Regressor
    - Adaboost Regressor
    - Decision Tree

3.1) Linear Regression

Algorithm Accuracy: Approximately 100% (Dynamic)



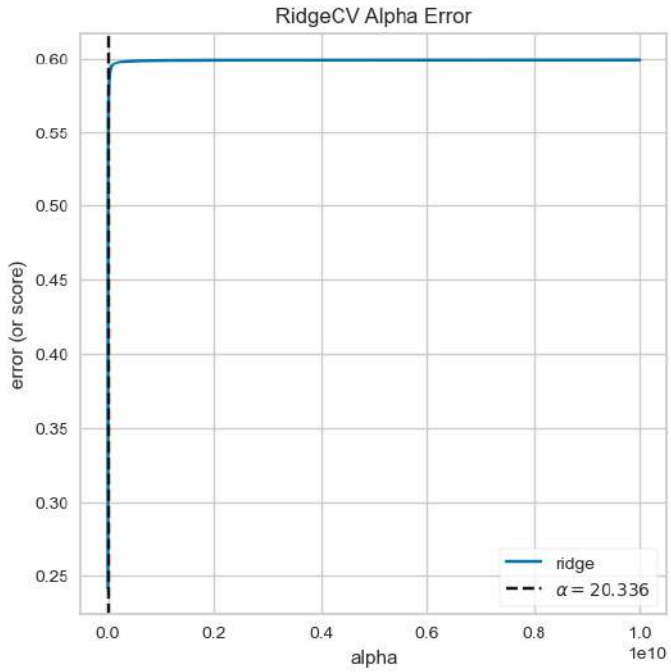
Linear Regression Model



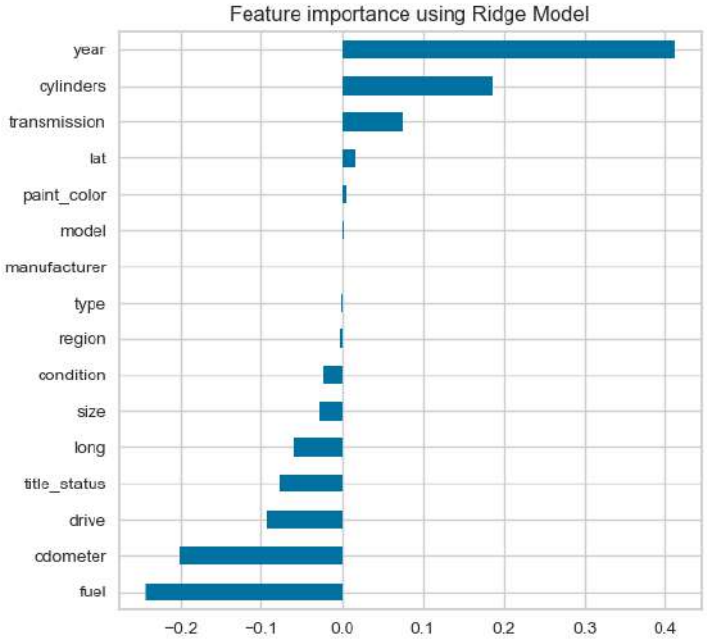
Performance Graph for Linear Regression Model

3.2) Ridge Regression

Algorithm Accuracy: 59.3%



Error graph of Ridge Regression Model



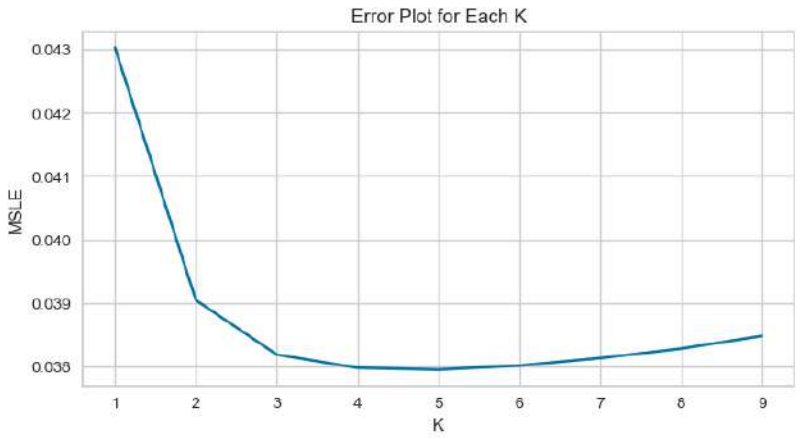
Performance Graph for Ridge Regression Model

3.3) Lasso Regression

Algorithm Accuracy:59.3%

3.4) KNN

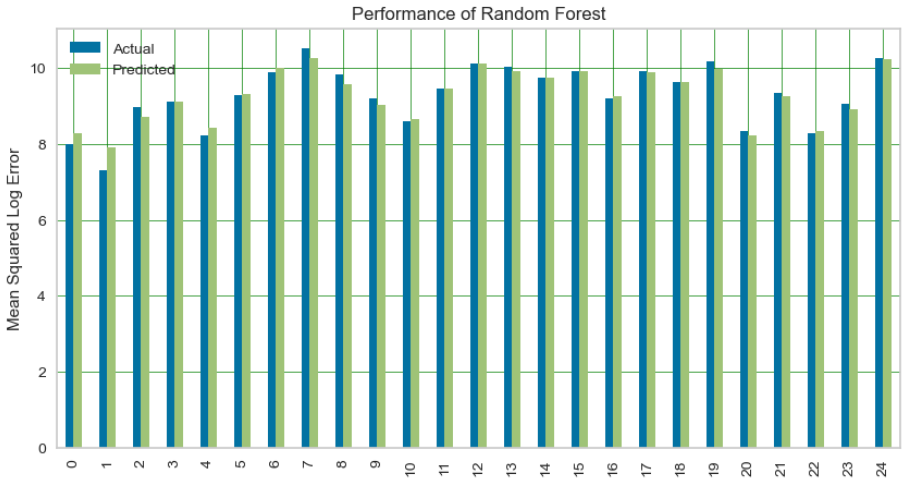
Algorithm Accuracy: 77%



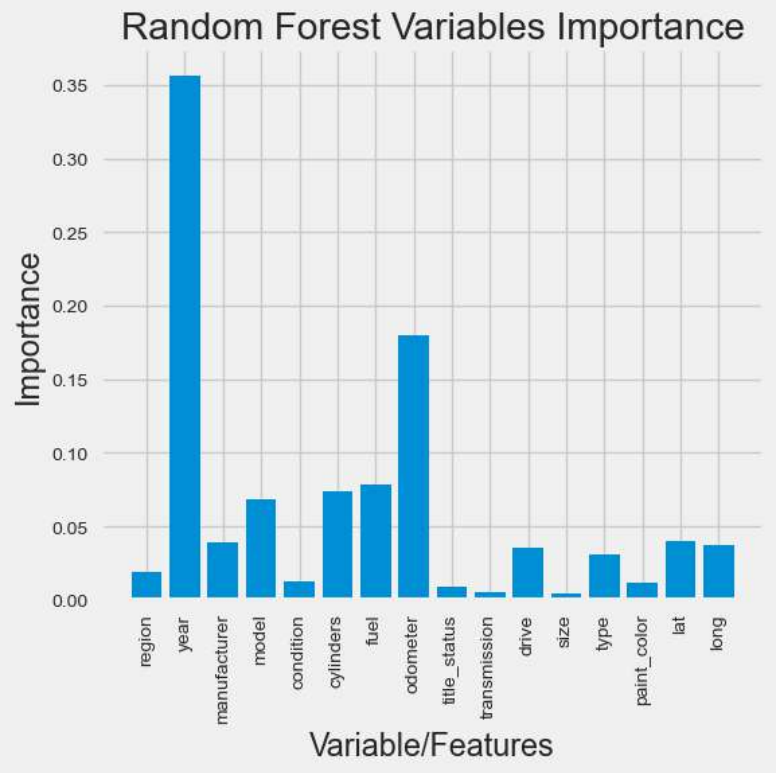
Error Graph for KNN Model

3.5) Random Forest

Algorithm Accuracy: Dynamic



Performance graph for Random Forest Model



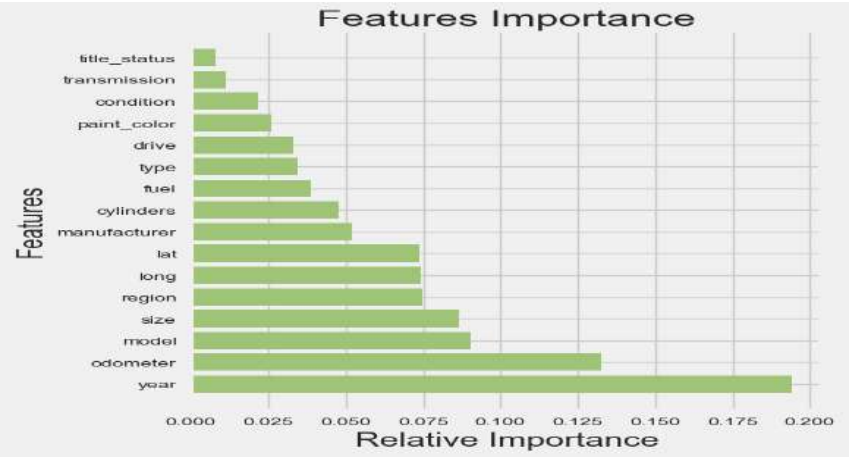
Random Forest Model

3.6) Bagging Regressor

Algorithm Accuracy: 77%

3.7) AdaBoost

Algorithm Accuracy: 87%



AdaBoost Model

3.8) Decision Tree

Algorithm Accuracy: Dynamic

Factors affecting prediction:

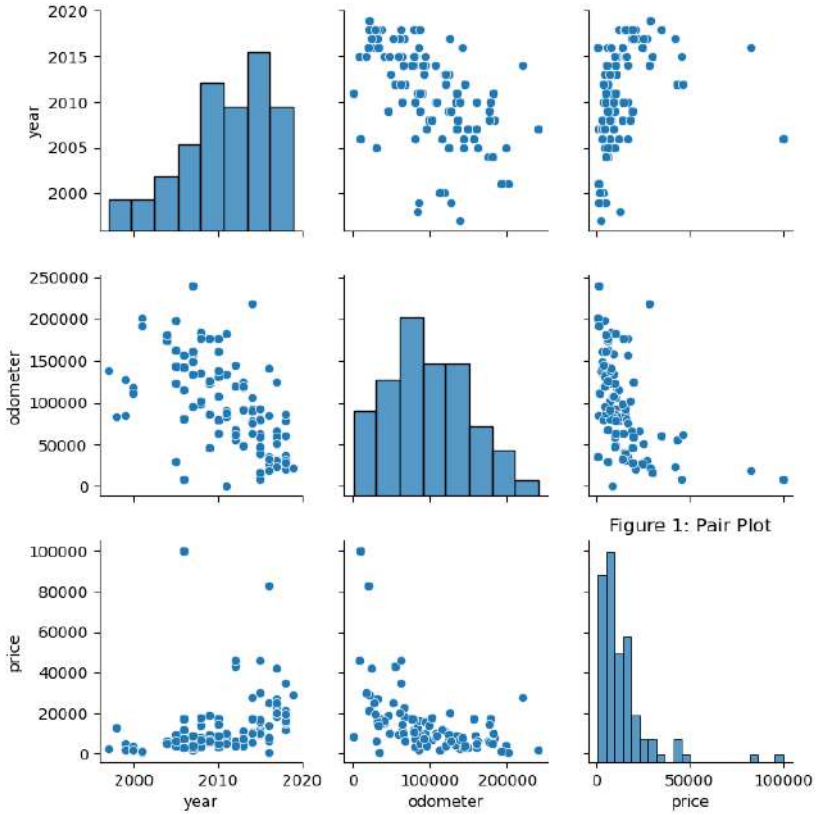
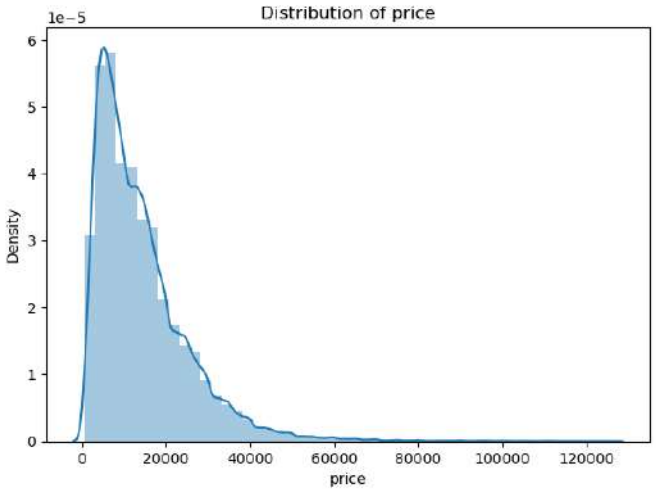
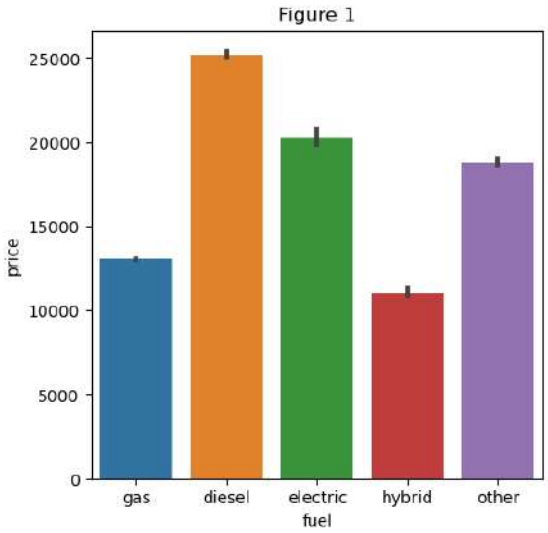


Figure 1: Pair Plot

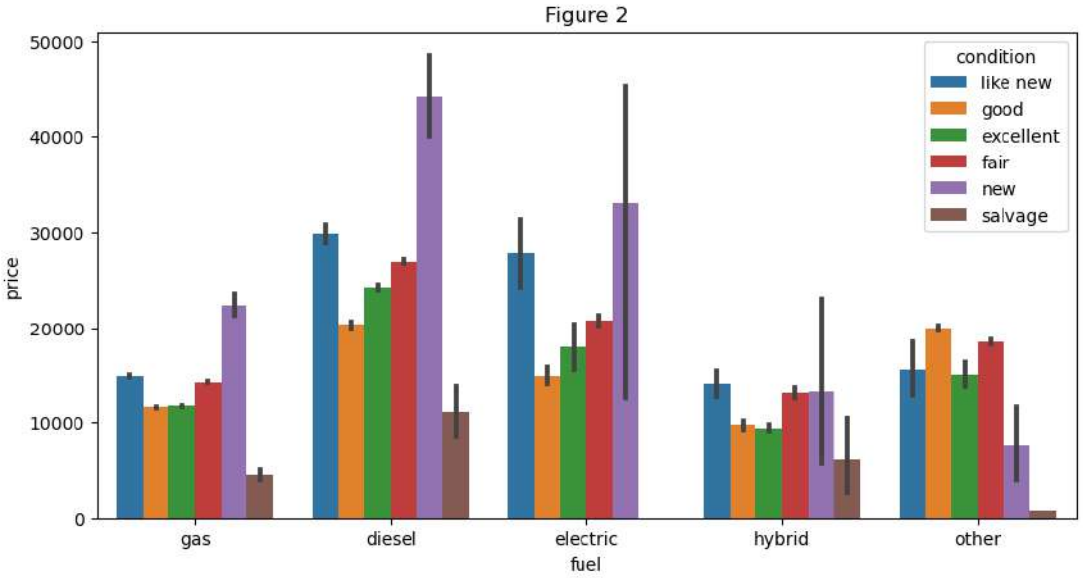
Factors to consider in price prediction of cars



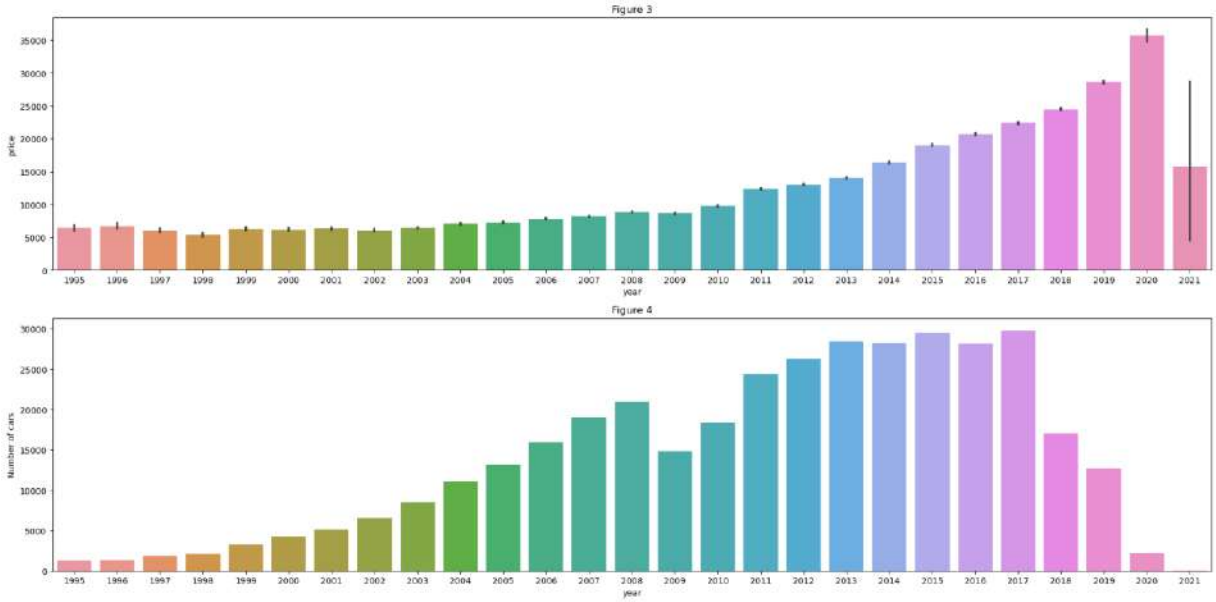
Distribution of price against density



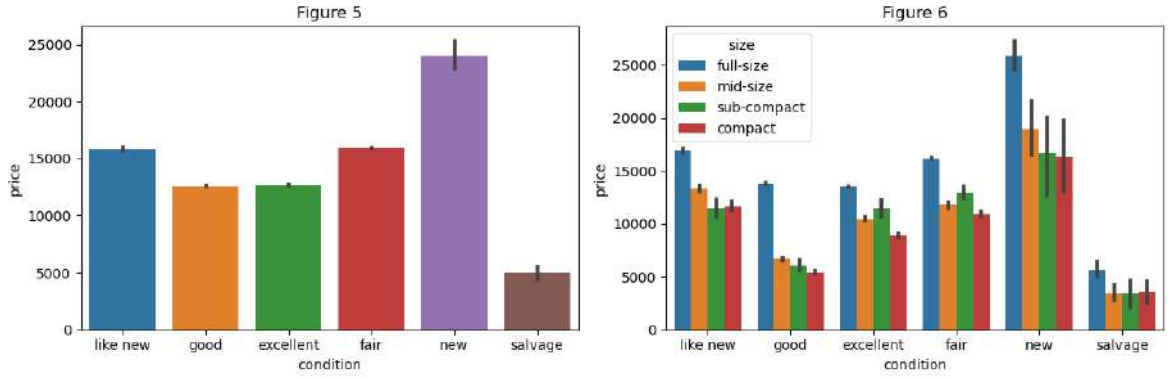
Types of fuels against price



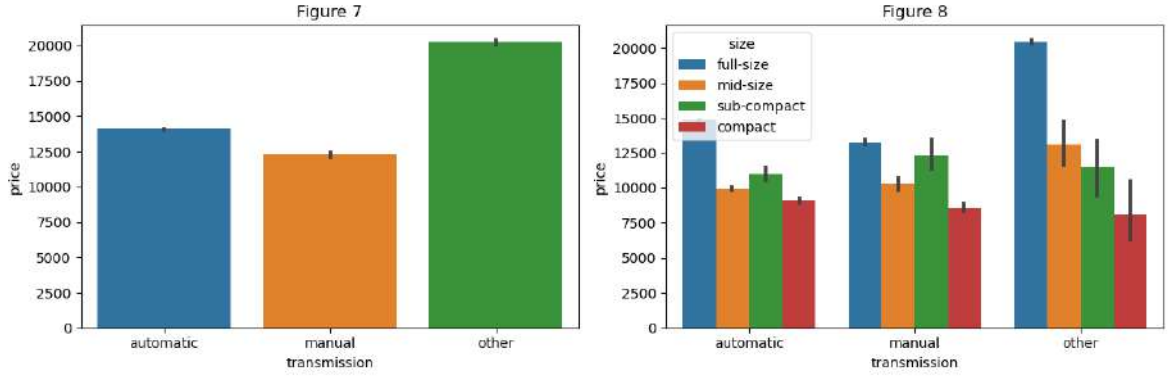
Types of fuels and condition against price



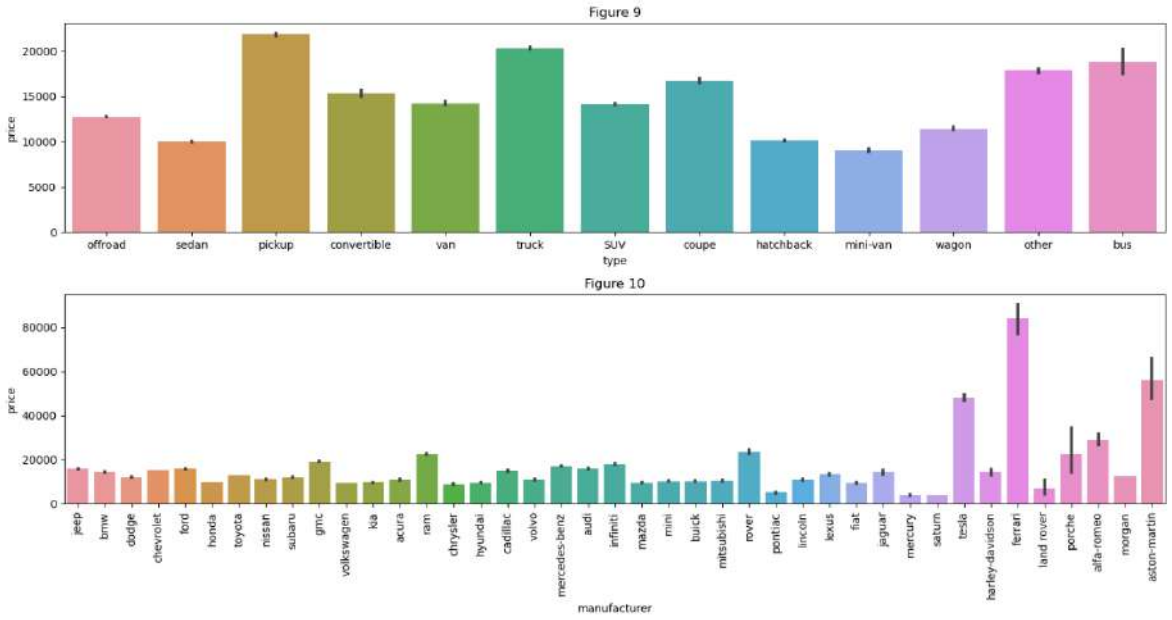
Price and number of cars against years



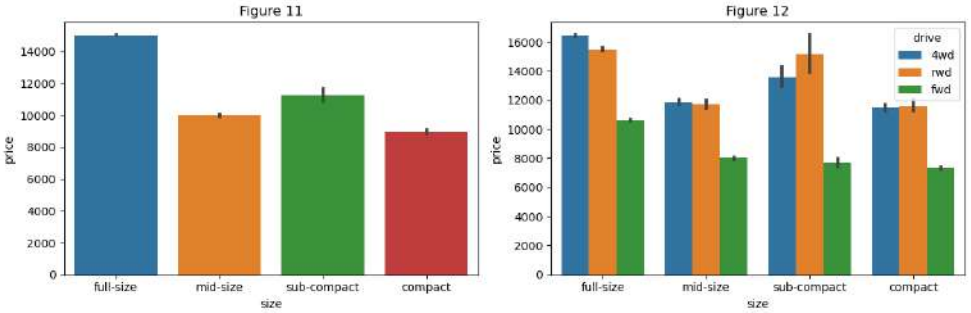
Condition and size against price



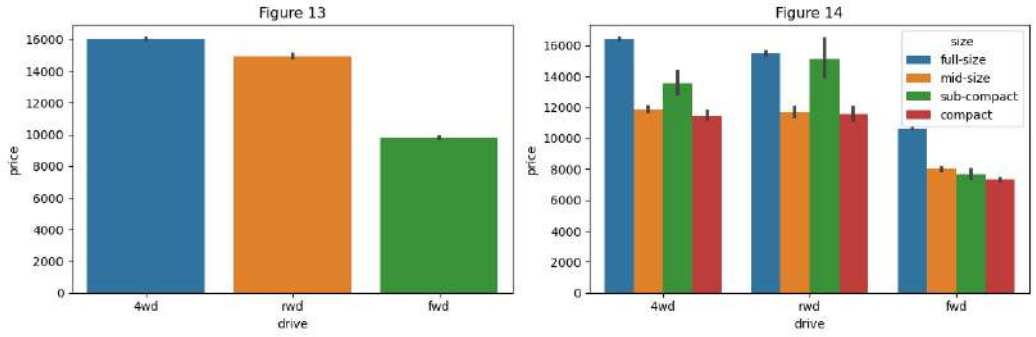
Transmission and size against price



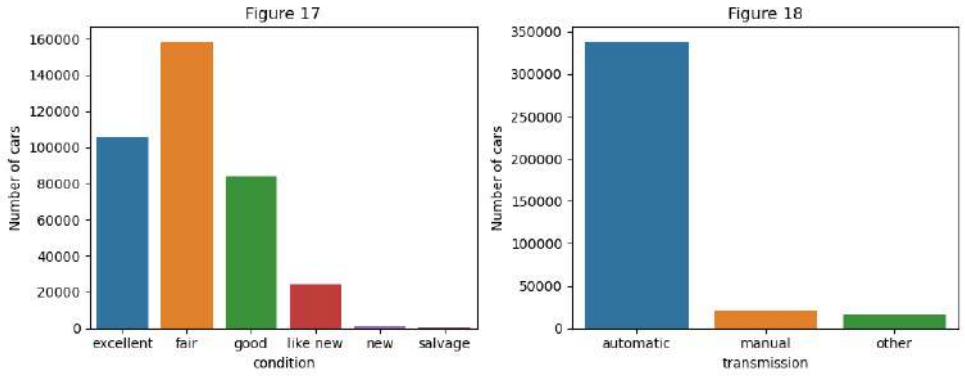
Manufacturer and type against price



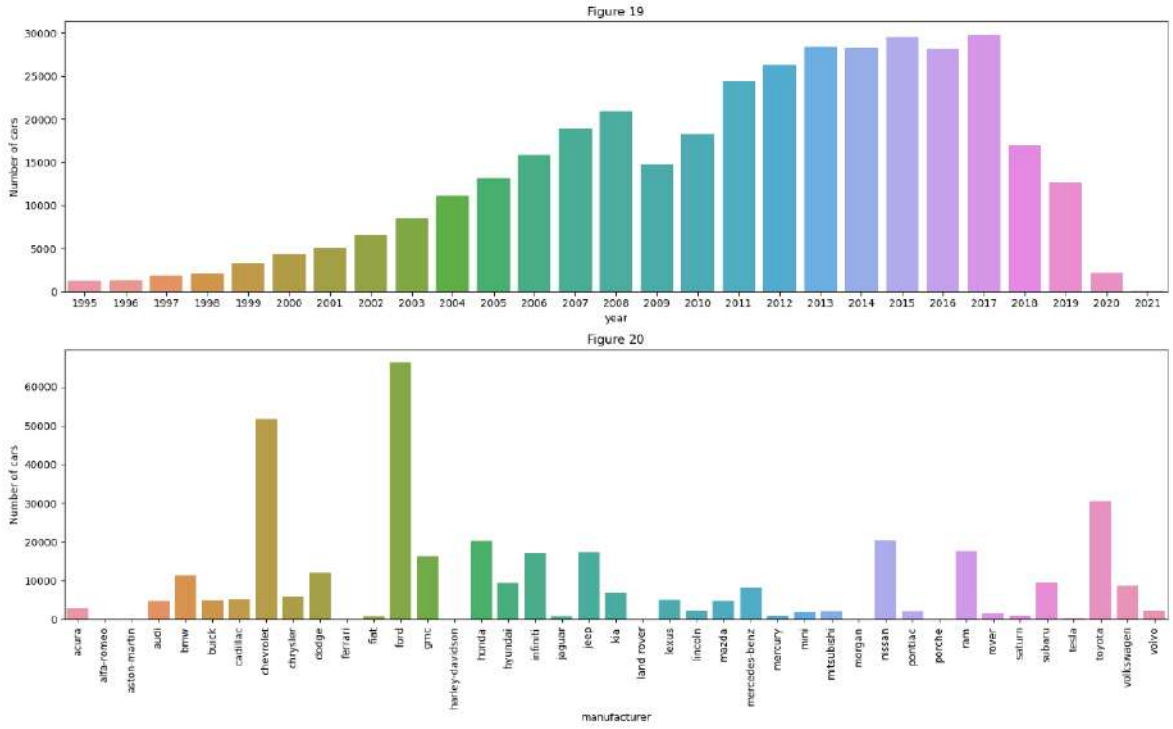
Size and drive against price



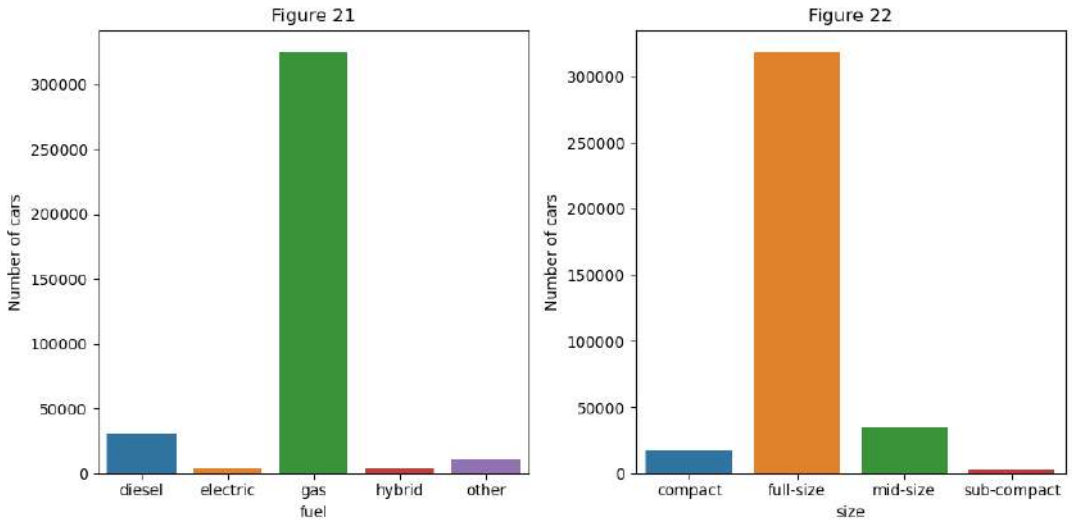
Drive and size against price



Condition and manual transmission against number of cars



Years and manufacturer against number of cars



Fuel and size against number of cars

Libraries used for Data Pre-Processing:

from sklearn.experimental import enable_iterative_imputer

from sklearn.impute import IterativeImputer

from sklearn.ensemble import ExtraTreesRegressor

from sklearn.linear_model import BayesianRidge

from sklearn.tree import DecisionTreeRegressor

from sklearn.neighbors import KNeighborsRegressor

from sklearn.preprocessing import OrdinalEncoder

from sklearn import preprocessing

from sklearn.model_selection import cross_val_score

from sklearn.impute import SimpleImputer

from sklearn.pipeline import make_pipeline

Steps:

1) Removed Irrelevant features.
2) Dealing with missing values.

    2.1) Estimating better iterative imputer method.

    2.2) Filling Numerical Missing Values i.e., year & odometer.

    2.3) Filling Categorical values.

3) Outliers

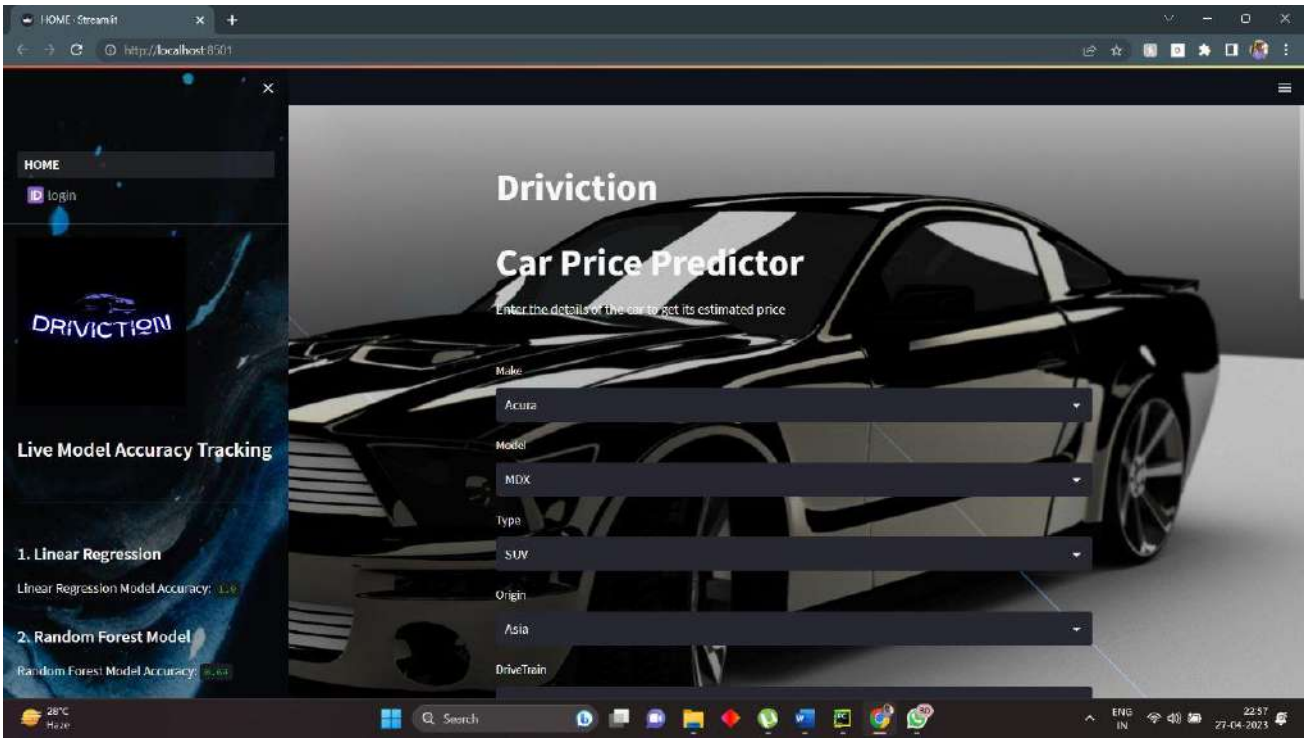
    3.1) Price

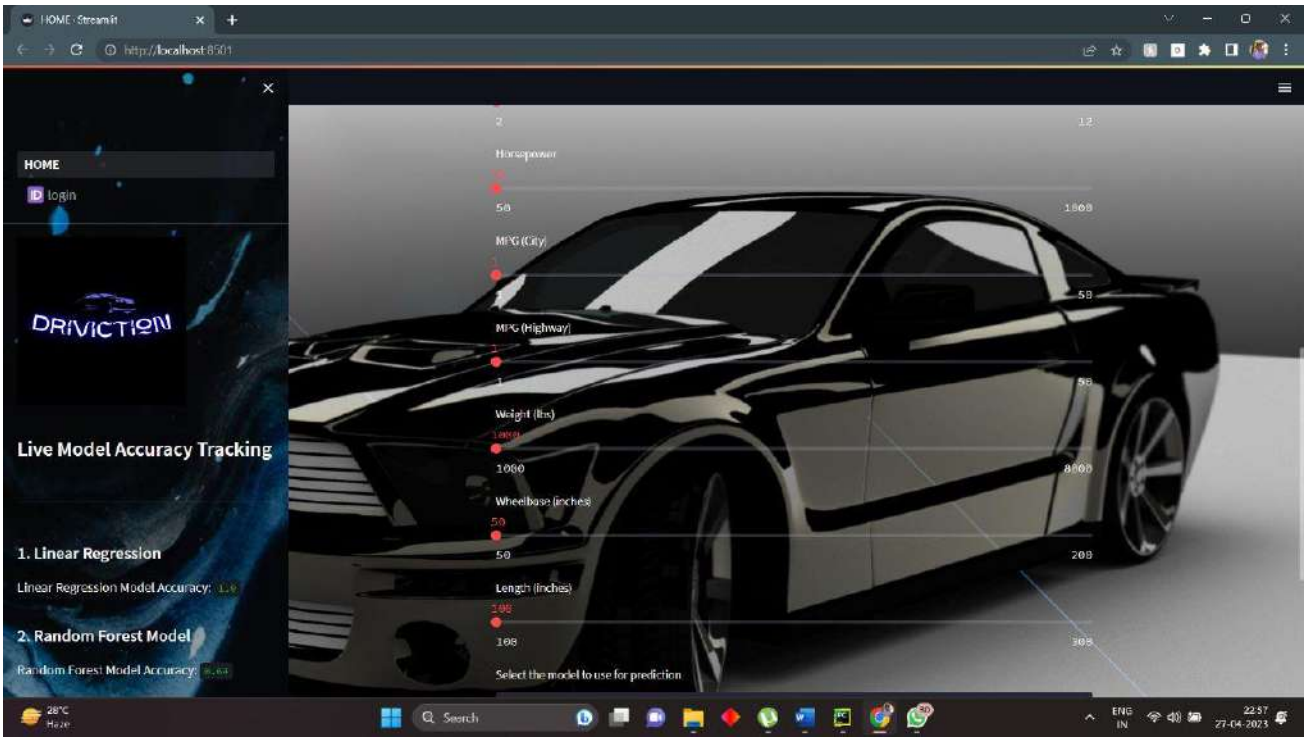
    3.2) Odometer

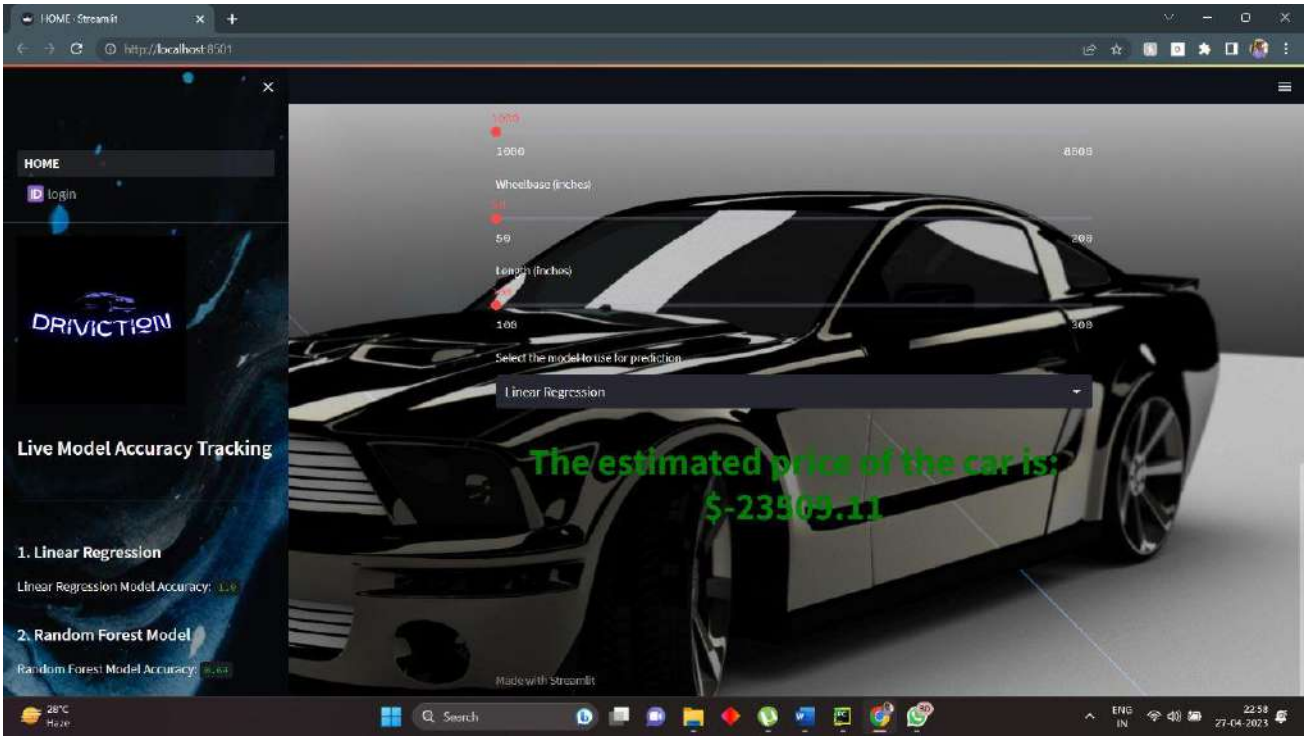
    3.3) Year

Chapter 6

Project Design

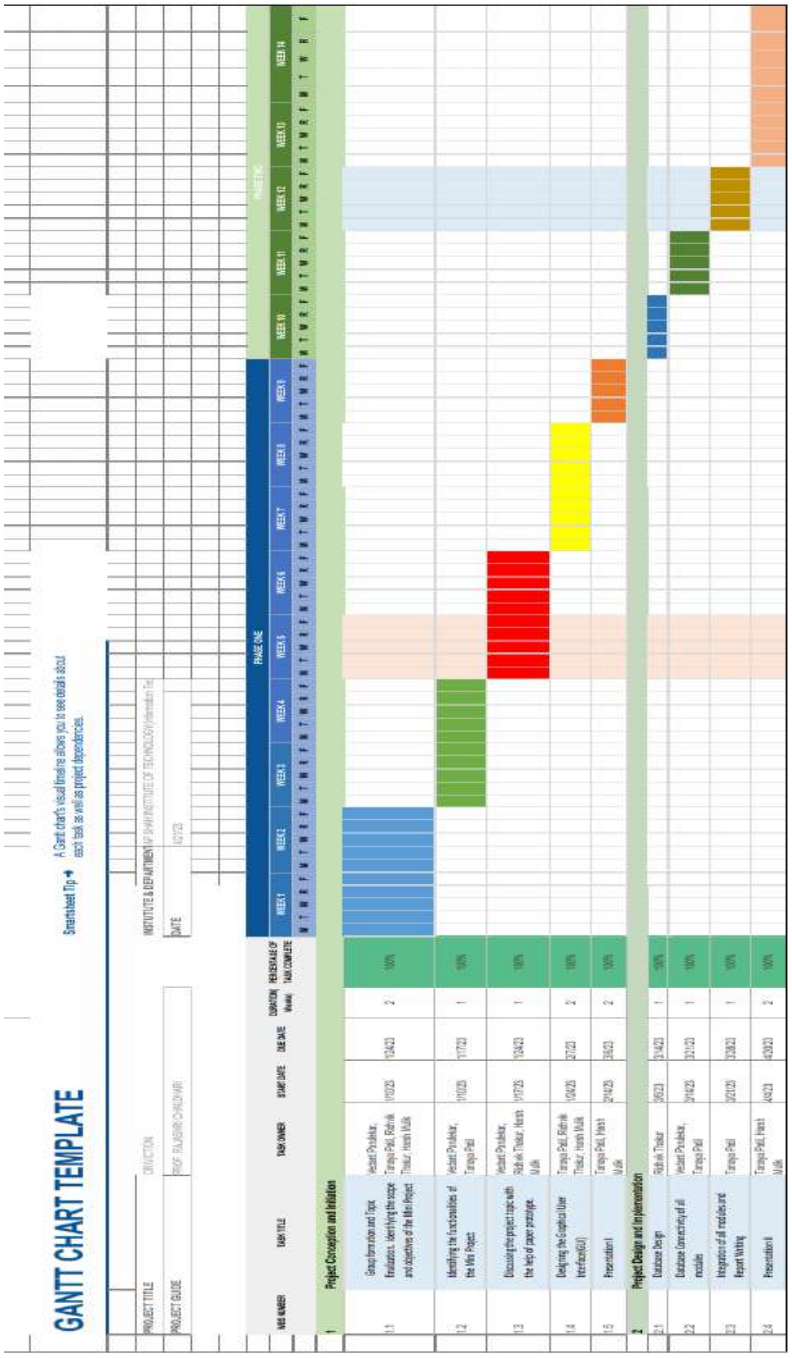





HOME · Streamlit
http://localhost:8501
HOME
login
DRIVICTION
Live Model Accuracy Tracking
1. Linear Regression
Linear Regression Model Accuracy:
2. Random Forest Model
Random Forest Model Accuracy:
1000
8500
Wheelbase (inches)
59
208
Length (inches)
100
300
Select the model to use for prediction
Linear Regression
The estimated price of the car is:
$-23509.11
Made with Streamlit
28°C
Haze
Search
ENG
IN
22:58
27-04-2023

Chapter 7

Project Scheduling



Chapter 8

Conclusion

By performing different Machine learning models, we aim to get a better result or one with less errors and with maximum accuracy. Our purpose was to predict the price of cars having 14 predictors and 429 data entries.

Initially, data cleaning is performed to remove the null values and outliers from the dataset then Machine Learning models are implemented to predict the price of cars.

Next, with the help of data visualization features were explored deeply. The relationship between the features is examined.

This is the conclusion of the machine learning model from creation to deployment.

Thus, concluded that Linear Regression is the best model for prediction of car prices for the given data set.